The Design Principle of Hash Function with Merkle-Damgård Construction

Duo Lei[1], Da Lin[2], Li Chao[1], Keqin Feng[3], and Longjiang Qu[1]

[1] Department of Science, National University of Defense Technology, Changsha, China
Duoduolei@163.com
[2] College of Mechanical and Electronic Control Engineering, Beijing Jiaotong University, BeiJing, China
[3] Department of Math, Tsinghua University, Beijing, China

Abstract. The paper discusses the security of hash function with Merkle-Damgård construction and provides the complexity bound of finding a collision and primage of hash function based on the condition probability of compression function $y = F(x, k)$. we make a conclusion that in Merkle-Dammaård construction, the requirement of free start collision resistant and free start collision resistant on compression function is not necessary and it is enough if the compression function with properties of fix start collision resistant and fix start preimage resistant. However, the condition probability $P_{Y|X=x}(y)$ and $P_{Y|K=k}(y)$ of compression function $y = F(x, k)$ have much influence on the security of the hash function. The best design of compression function should have properties of that $P_{Y|X=x}(y)$ and $P_{Y|K=k}(y)$ are both uniformly distributed for all x and k. At the end of the paper, we discussed the block cipher based hash function, point out among the the 20 schemes, selected by PGV[2] and BPS[12], the best scheme is block cipher itself, if the block cipher with perfect security and perfect key distribution.

1 Introduction

Most of hash functions iterated a compression function by Merkle-Damgård construction with constant IV[3]. A well known approach for building hash function is the compression function out of a block cipher which have been discussed sine Rabin[10] given the first model of that kind of structure. As BRS point out the block cipher approach has been less widely used for variety of reasons, and the emergence of the AES[13] has somewhat modified this landscape, especially recently the MD5 and SHA1 are attacked[8][9][14][16].

The topics of building hash function based on block cipher have been systematically analyzed in paper [2] [4][7][12][15]. The PGV paper considered turning a block cipher $E : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^n$ into a hash function $H : (\{0,1\}^n)^* \to \{0,1\}^n$ using a compression function $F : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^n$ derived from E. For v is a fixed n-bit constant, PGV considers all 64 compression functions

F of the form $F(h_{i-1}, m_i) = E_a(b) \oplus c$ where $a, b, c \in \{h_{i-1}, m_i, h_{i-1} \oplus m_i, v\}$, then define the iterated hash of F as $H(M_1, \ldots, M_t, IV) = h_t, h_0 = IV, h_i = F(h_{i-1}, M_i), i \in [1..t]$, $|M_i| = n$. Of the 64 such schemes, the authors of [2] regarded 12 as secure. In fact, PGV[2] implied the condition of free start not fix start. The authors of [12] taken a more proof-centric look at the schemes from PGV, proved additional 8 schemes were collision resistant, divided the 20 schemes into two group where the $group-1$ were the 12 schemes picked by PGV and the $group-2$ were the new founded 8 schemes. For the new founded schemes, the hash function H immune to collision attack within the Merkle-Damgård paradigm, the compression functions were not immune to collision attack, the proves of collision resistant of $group-2$ used the assumptions of E with black box model and H with fix start model. They also provided both upper and lower bounds for each scheme.

This paper takes a proof-centric look at the schemes based on the probability theory, providing the exact probability of finding a collision or primage based on the assumption of known condition probability of block cipher E. Let $F : \{0,1\}^{\iota} \times \{0,1\}^{\kappa} \to \{0,1\}^n$, $x \in \{0,1\}^{\iota}$, $y \in \{0,1\}^n$,$k \in \{0,1\}^{\kappa}$ and Hash function H is Merkle-Damgård construction hash function with compression function F.

Firstly, the probability of finding a collision or a preimage is defined by complexity of finding collision in one time computation of compression function F, based on which a more precise definitions of free start and fix start of collision resistant and preimage resistant are given about compression function and Hash function. In our point of view, if we have no way to find the collision or preimage except exhaustive search then the function is called collision resistant or preimage resistant. Secondly, the upper bound probabilities of finding collision and preimage about the compression function F and the Hash function with M-D paradigm are given which is based on the condition probability of $P_{Y|X=x}(y)$ and $P_{Y|K=k}(y)$. At last we analyze the 64 schemes with M-D construction, in fact the best compression function to build M-D hash function is block cipher E itself, the best block cipher is the block cipher designed with perfect security and perfect key distribution.

The paper is organized as follows. The mathematical preliminaries and notation employed are described in section 2. The preimage resistance and collision resistance of compression function are given in section 3,that of a hash function are presented in section 4. Section 5 describe the preimage resistance and collision resistance of PGV schemes. Section 6 is our conclusion.

2 Definition

2.1 Basic Definition

A discrete random variable X is a mapping from the sample space Ω to an alphabed $\mathcal{X}$. X assigns a value $x \in \mathcal{X}$ to each elementary event in the Ω and the probability distribution of X is the function[5]

$$P_X : \mathcal{X} \to \Re : x \mapsto P_X(x) = P[X = x] = \sum_{\omega \in \Omega : X(\omega) = x} P[\omega].$$

If the conditioning event involves another random variable Y defined on the same sample space, the conditional probability distribution of X given that Y takes on a value y is:

$$P_{X|Y=y}(x) = \frac{P_{XY}(x,y)}{P_Y(y)}$$

whenever $P_Y(y)$ is positive . Two random variables X and Y are called independent if for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$:

$$P_{XY}(x,y) = P_X(x) \cdot P_Y(y).$$

Let $F : \{0,1\}^\iota \times \{0,1\}^\kappa \to \{0,1\}^n$, $x \in \{0,1\}^\iota$, $y \in \{0,1\}^n$,$k \in \{0,1\}^\kappa$ and $y = F_k(x)$. If $F : \{0,1\}^\iota \times \{0,1\}^\kappa \to \{0,1\}^n$ is a compression function of hash function H_X and H_K:

$$H_X : \{0,1\}^n \times \{0,1\}^{\kappa \cdot *} \to \{0,1\}^n, x \in \{0,1\}^n, m \in \{0,1\}^{\kappa \cdot *}, y \in \{0,1\}^n$$

$$z = H_X(m,x) \stackrel{\triangle}{=} H_X(m_t \| \ldots \| m_1, x) = F_{m_t}(F_{m_{t-1}}(\ldots(F_{m_1}(x))\ldots))$$

$$H_K : \{0,1\}^{\iota \cdot *} \times \{0,1\}^n \to \{0,1\}^n, k \in \{0,1\}^n, m \in \{0,1\}^{\iota \cdot *}, y \in \{0,1\}^n$$

$$z = H_K(m,k) \stackrel{\triangle}{=} H_K(m_t \| \ldots \| m_1, k) = F_{\ldots F_{F_k(m_1)}(m_2)\ldots}(m_t)$$

$$H_K(m_2 \| m_1, k) = F_{F_k(m_1)}(m_2).$$

In this paper, if no special statement are given, the function H_K, H_X and F are defined as above and x, k are uniformly distributed in definition domain. q_k,q_x and q_y denote $q_k \stackrel{\triangle}{=} \max_y P_{Y|K=k}(y)2^\iota$, $q_x \stackrel{\triangle}{=} \max_y P_{Y|X=x}(y)2^\kappa$ and $q_y \stackrel{\triangle}{=} P_Y(y)2^\iota 2^\kappa$.

In hash function attack, the probability of finding a primage or collision is different from tradition point of view of probability. If the compression function F is block cipher E with form of $E_k(x) = y$, then the probabilities of $P_{X|Y=y,K=k}(x)$, $P_{K|Y=y,X=x}(k)$ are both equal 0 or 1 (assume the cipher with perfect key distribution). However, for given y, k the value x satisfying $y = E_k(x)$ can be found directly by computing $x = E_k^{-1}(y)$, but for given y, x the value k satisfying $y = E_k(x)$ can be found only by exhaustive search of k, that implies we should compute E for each guessing k. So we consider giving new definition about the probability of finding collision or preimage based on the times computing F being made.

Definition 1. *Let X,Y and K are random variables defined in $\mathcal{X}$,$\mathcal{K}$ and $\mathcal{Y}$, respectively, Y can be deduced from X, K by $y = F_k(x)$, E be a set with variables of X, K, Y, the Q_E is defined as the max probability of finding the inner relations of variables in E with one time computation of function F.*

More precisely, Q_{XYK} means the max probability of finding values x, y, k satisfying $y = F_k(x)$ in one time computation of F, in fact for any x, k, we can compute y by $F_k(x)$, which gives $Q_{XKY} = 1$. $Q_{XX'K}$ means the max probability

of finding x, x' satisfying $F_k(x) = F_k(x')$ in one time computation of F, if $\Lambda \in \mathcal{X}$, let $Q_{XX'_\Gamma K}$ means the max probability of finding $x, x' \in \Gamma$ satisfying $F_k(x) = F_k(x')$ in one time computing of F. The condition probability $Q_{XK|Y=y}$ implies for given $y \in \mathcal{Y}$ the max probability of finding x, k satisfying $y = F_k(x)$ in one time computation of F, and $Q_{XK|Y} = \max_y Q_{XK|Y=y}$. If p is the max probability of finding the inner relation of E in t times computation of F, then $Q_E = \frac{p}{t}$, where if F is invertible, the one times of computing F^{-1} is regarded as one times computation of F.

If F is block cipher, from given y, k we can compute $x = F_k^{-1}(y)$ that means $Q_{XK|Y=y} = Q_{X|K=k,Y=y} = 1$. But we can't compute k directly from give y, x, the only way to find k is exhaustive search, we have $Q_{K|X=x,Y=y} \geq P_{Y|X=x}(y)$.

Definition 2 (Perfect Secrecy[6]). *A cryptosystem has perfect secrecy if*

$$P_{X|Y=y}(x) = P_X(x)$$

for all $x \in \{0,1\}^n, y \in \{0,1\}^n$.

Definition 3 (Perfect Key Distribution). *A cryptosystem has perfect key distribution if*

$$P_{K|Y=y}(k) = P_K(k)$$

for all $x \in \{0,1\}^n, y \in \{0,1\}^n$.

In fact, $P_{XY}(xy) = P_{X|Y=y}P_Y(y) = P_{Y|X=x}(y)P_X(x)$, since $P_{X|Y=y}(x) = P_X(x)$, we get $P_{Y|X=x}(y) = P_Y(y)$.

Definition 4 (Random Oracles[12]). *A fixed-size random oracle is a function* $f : \{0,1\}^\iota \to \{0,1\}^n$*, chosen uniformly at random from the set of all such functions. For interesting sizes* a *and* b*, it is infeasible to implement such a function, or to store its truth table. Thus, we assume a public oracle which, given* $x \in \{0,1\}^\iota$*, computes* $y = f(x) \in \{0,1\}^n$.

Definition 5 (Free Start Collision resistant). *We call* F, H_X *or* H_K *is collision resistant if there is no way to find collision except exhaustive search. And* $Q_{<XX'KK'>}$*,* $Q_{<XX'MM'>}$ *and* $Q_{<MM'KK'>}$ *denote the max probability of finding the collision of* F, H_X *and* H_K*, in one time computation of* F*, respectively.*

Definition 6 (Fix Start Collision resistant). *Let* $\Lambda \subset \{0,1\}^\iota$*,* $\Gamma \subset \{0,1\}^\kappa$*, we call* F *or* H_X *is collision resistant with fix start* X*, if there is no way to find collision* $F_k(x) = F_{k'}(x')$ *or* $H_X(m,x) = H_X(m',x')$ *except exhaustive search of* k *or* k'*, where* $x, x' \in \Lambda$*. And* F *or* H_K *is collision resistant with fix start* K*, if there is no way to find collision* $F_k(x) = F_{k'}(x')$ *or* $H_K(m,k) = H_K(m',k')$ *except exhaustive search, where* $k, k' \in \Lambda$*. And* $Q_{<XX'_\Lambda KK'>}$*,*$Q_{<XX'_\Lambda MM'>}$ *represent the max probabilities of finding the collision of* F *and* H_X *with fix start* X*, in one time computation of* F*, respectively.* $Q_{<XX'KK'_\Gamma>}$*,*$Q_{<MM'KK'_\Gamma>}$ *represent the max probabilities of finding the collision of* F *and* H_K *with fix start* K*, in one time computation of* F*, respectively.*

Definition 7 (Free Start Preimage Resistant). *We call F(H_X or H_K) is preimage resistant if, for given y, no way to find (x,k)((x,m) or (k,m)) satisfying $y = F_k(x)$($y = H_X(m,x)$ or $y = H_K(m,k)$) except exhaustive search. And $Q_{[XK|Y]}$, $Q_{[XM|Y]}$ and $Q_{[MK|Y]}$ denote the max probabilities of finding the preimage of F,H_X and H_K, for some given y in one time computation of F, respectively.*

Definition 8 (Fix Start Preimage Resistant). *Let $\Lambda \subset \{0,1\}^\iota$, $\Gamma \subset \{0,1\}^\kappa$, we call F or H_X is preimage resistant with fix start X, if for given y, no way to find (k, x) or (m,x) satisfying $y = F_k(x)$ or $y = H_X(m,x)$, where $x \in \Lambda$, except exhaustive search of k, we call F or H_K is primage resistant for fix start K, if for given y no say to find (k,x) or (k,m) satisfying $y = F_k(x)$ or $H_K(m,k)$ expect exhaustive search of m where $k \in \Gamma$. And $P_{[X_\Lambda K|Y]}$ denotes the max probability of finding the x,k satisfying $y = F_k(x)$ in one time computation of F for random selected y and $x \in \Lambda$. $Q_{[X_\Lambda K|Y]}$ denotes the max probability of finding the x,k satisfying $y = F_k(x)$ for random selected y in one time computation where $k \in \Gamma$.*

3 Hash Properties of Compression Function

The conclusions of this section are that the best design of F should satisfy y is uniformly distributed in $\{0,1\}^n$ for each $k \in \{0,1\}^\kappa$ and for each $x \in \{0,1\}^\iota$, no matter for free start or fix start and for preimage resistance or collision resistance.

3.1 Free Start Preimage Resistance

The conclusion of this subsection is Theorem1, the upper bound of free start preimage resistant of F is $\max_{x,k}\{q_k 2^{-\iota}, q_x 2^{-\kappa}\}$, which implies the best selection of free start collision resistant and free start preimage resistant have same requirement on F.

Lemma 1. $Q_{[XK|Y]} = \max_y\{Q_{XKY=y}, Q_{X|KY}, Q_{K|XY}\}$.

Theorem 1. *If F is free start Preimage resistance then*

$$Q_{[XK|Y]} \leq \max_{x,k}\{q_k 2^{-\iota}, q_x 2^{-\kappa}\}. \tag{1}$$

Proof. Given y,k finding $x_1,\ldots,x_t$ with $y = F_k(x_i)$ the success probability is:

$$p = 1 - \prod_{i=0}^{t-1} \frac{(2^\iota - 2^\iota P_{Y|K=k}(y) - i)}{(2^\iota)(2^\iota - 1)\ldots(2^\iota - t + 1)}$$

Let denote $n = 2^\iota$ then

$$p = 1 - \prod_{i=0}^{t-1}(1 - \frac{q_k}{n-i}) \approx 1 - \prod_{i=0}^{t-1} \exp^{\frac{q_k}{n-i}} \approx 1 - \prod_{i=0}^{t-1} \exp^{(\frac{q_k}{n} + \frac{iq_k}{n^2})}$$

We get $Q_{X|YK} \leq q_k 2^{-\iota}$, similarly we get $Q_{K|YX} \leq q_x 2^{-\kappa}$. □

3.2 Free Start Collision Resistance

Conclusion of this subsection is Theorem2, upper bound of free start collision resistant of F is smaller than $\max_{x,k,y}\{\sqrt{(q_x-1)2^{-\kappa}}, \sqrt{(q_k-1)2^{-\iota}}, \sqrt{(q_y-1)2^{-\iota-\kappa}}\}$, which implies the best design of F should satisfy y is uniformly distributed in $\{0,1\}^n$ for each $k \in \{0,1\}^\kappa$ and for each $x \in \{0,1\}^\iota$.

Lemma 2. $Q_{<XX'KK'>} = \max\{Q_{XX'K}, Q_{XKK'}, Q_{XX'KK'}\}$

Proof. If f is collision resistant, getting a collision of F has three ways, firstly, searching x, x', k satisfying $F_k(x) = F_k(x')$, secondly searching x, k, k' satisfying $F_k(x) = F_{k'}(x)$, and thirdly, searching x, x', k, k' satisfying $F_k(x) = F_{k'}(x')$. □

Lemma 3. *F is collision resistant, then*

- $Q_{XX'K} \le \max_k \sqrt{(q_k-1)2^{-\iota}}$
- $Q_{XKK'} \le \max_{x,y} \sqrt{(q_x-1)2^{-\kappa}}$
- $Q_{XX'KK'} \le \max_y \sqrt{(q_y-1)2^{-\iota-\kappa}}$

Proof. F is collision resistant, the collision can be get only by exhaustive search.

- The fastest way to search for collision is the way based on birthday paradox. For random selected k searching $x_1, x_2, \ldots x_t$ finding collision of $F_k(x_i) = F_k(x_j)$. The max probability of success is

$$p = 1 - \frac{2^\iota(2^\iota - 2^\iota P_{Y|K=k}(y_1))\ldots(2^\iota - \sum_i^{t-1}(2^\iota P_{Y|K=k}(y_i))}{\binom{2^\iota}{t} t!}$$

  Let denote $n \triangleq 2^\iota$ and $q_k \triangleq 2^\iota \max_y P_{Y|K=k}(y)$ then

$$\begin{aligned} p &\le 1 - \frac{(n)(n-q_k)\ldots(n-q_k(t-1))}{(n)(n-1)\ldots(n-t+1)} \\ &= 1 - \prod_{i=0}^{t-1}\frac{n-iq_k}{n-i} = 1 - \prod_{i=0}^{t-1}(1-\frac{iq_k-i}{n-i}) = 1 - \prod_{i=0}^{t-1}(1-\frac{i}{n-i}(q_k-1)) \\ &\approx 1 - \prod_{i=0}^{t-1}\exp^{\frac{i}{n-i}(q_k-1)} \approx 1 - \prod_{i=0}^{t-1}\exp^{(\frac{i}{n}+\frac{i^2}{n^2})(q_k-1)} \end{aligned}$$

  Same as birthday paradox, when $t \ge \sqrt{n/(q_k-1)}$, $q_k > 1$ the success probability of collision is bigger than 1/2. We get $Q_{XX'|K} \le \sqrt{\frac{q_k-1}{2^\iota}}$.
- similar as item 1, we get $Q_{KK'|X} \le \sqrt{\frac{q_x-1}{2^\kappa}}$;

– similar as item 1, we get $Q_{KK'XX} \leq \sqrt{\frac{q_y-1}{2^{\kappa+\iota}}}$. □

Theorem 2. *F is collision resistant then*

$$Q_{<XX'KK'>} \leq \max_{x,k,y}\{\sqrt{(q_x-1)2^{-\kappa}}, \sqrt{(q_k-1)2^{-\iota}}, \sqrt{(q_y-1)2^{-\iota-\kappa}}\} \quad (2)$$

3.3 Fix Start Preimage Resistance

The conclusions of this subsection are Theorem3 and Theorem4.

Lemma 4. *Let $\Lambda \subset \{0,1\}^{\iota}$, $\Gamma \subset \{0,1\}^{\kappa}$, P_{X_Λ}, P_{K_Γ} to denote the probability of $x \in \Lambda$*

– $Q_{X_\Lambda K|Y=y} = \max\{\max_{x\in\Gamma} Q_{K|Y=y,X=x}\ , Q_{XK|Y=y}P_{X_\Lambda}\}$
– $Q_{XK_\Gamma|Y=y} = \max\{\max_{k\in\Gamma} Q_{X|Y=y,K=k}\ , Q_{XK|Y=y}P_{K_\Gamma}\}$

Proof. For $Q_{X_\Lambda K|Y=y}$, the preimage can be found in two ways, firstly for selected $x \in \Gamma$, find the preimage of y; secondly, find the preimage of y, then check $x \in \Lambda$ being satisfied or not. □

Lemma 5. *If $Q_{X|Y=y,K=k} = 1 \wedge Q_{K|Y=y,X=x} = q_x 2^{-\kappa}$ then:*

– $Q_{X|Y=y,K=k:k,\in\Gamma} = 1$
– $Q_{K|Y=y,X=x:x\in\Lambda} \leq q_x 2^{-\kappa}$

Lemma 6. *If $Q_{K|Y=y,X=x} = 1 \wedge Q_{X|Y=y,K=k} = q_k 2^{-\iota}$ then:*

– $Q_{K|Y=y,X=x:x\in\Lambda} = 1$
– $Q_{X|Y=y,K=k:k\in\Gamma} \leq q_k 2^{-\iota}$

Theorem 3. *If If $Q_{X|Y=y,K=k} = 1 \wedge Q_{K|Y=y,X=x} = q_x 2^{-\kappa}$ then*

$$Q_{[XK_\Gamma|Y]} = 1. \quad (3)$$

$$Q_{[X_\Lambda K|Y]} \leq \sum_{x\in\Lambda} q_x 2^{-\kappa} \quad (4)$$

Proof. The Eq.(3) can be get directly from Lemma5.

$$Q_{K|Y=y,X=x} = q_x 2^{-\kappa} \Rightarrow Q_{K|Y=y,X=x:x\in\Lambda} \leq q_x 2^{-\kappa}$$

$$Q_{X|Y=y,K=k} = 1 \Rightarrow Q_{X=x_0,K|Y=y} = P_{Y|X=x_0}(y)$$

From Lemma4 we get the conclusion. If $\forall k, k', k \neq k'$ and $\forall x, x' \in \Lambda$ we have $F_k(x) \neq F_{k'}(x')$, then the equation being hold.

Theorem 4. *If If $Q_{K|Y=y,X=x} = 1 \wedge Q_{X|Y=y,K=k} = q_k 2^{-\iota}$ then:*

$$Q_{[X_\Lambda K|Y]} = 1. \quad (5)$$

$$Q_{[XK_\Gamma|Y]} \leq \sum_{k\in\Gamma} q_k 2^{-\iota} \quad (6)$$

3.4 Fix Start Collision Resistance

The conclusion of this subsection are Theorem5 and Theorem6, which tell us the best design of F also should satisfy Y is uniformly distributed in $\{0,1\}^n$ for each $k \in \{0,1\}^\kappa$ and for each $x \in \{0,1\}^\iota$.

Lemma 7. *Let* $\Lambda \subset \{0,1\}^\iota$, $\Gamma \subset \{0,1\}^\kappa$

- $Q_{<XX'_\Lambda KK'>} = \max\{Q_{XX'_\Lambda K}, Q_{X_\Lambda KK'}, Q_{XX'_\Lambda KK'}\}$
- $Q_{<XX'KK'_\Gamma>} = \max\{Q_{XX'K_\Gamma}, Q_{XKK'_\Gamma}, Q_{XX'KK'_\Gamma}\}$.

Lemma 8. *Let* $\Lambda \subset \{0,1\}^\iota$

- $Q_{X_\Lambda KK'} = \max\limits_{x\in\Lambda}\{Q_{KK'|X=x}, P_{X_\Lambda}Q_{XKK'}\}$
- $Q_{XX'_\Lambda K} = \max\limits_{x,x'\in\Lambda}\{Q_{K|X=x,X'=x'}, P_{X_\Lambda}Q_{XK|X'=x'}, P_{X_\Lambda X'_\Lambda}Q_{XX'K}\}$
- $Q_{XX'_\Lambda KK'} = \max\limits_{x,x'\in\Lambda}\{Q_{KK'|X=x,X'=x'}, P_{X_\Lambda}Q_{XKK'|X'=x'}, P_{X_\Lambda X'_\Lambda}Q_{XX'KK'}\}$

Lemma 9. *Let* $\Gamma \subset \{0,1\}^\kappa$

- $Q_{XX'K_\Gamma} = \max\limits_{k\in\Gamma}\{Q_{XX'|K=k}, P_{K_\Gamma}Q_{XX'K}\}$
- $Q_{XKK'_\Gamma} = \max\limits_{k,k'\in\Gamma}\{Q_{X|K=k,K'=k'}, P_{K_\Gamma}Q_{XK|K'=k'}, P_{K_\Gamma K'_\Gamma}Q_{XKK'}\}$
- $Q_{XX'KK'_\Gamma} = \max\limits_{k,k'\in\Gamma}\{Q_{XX'|K=k,K'=k'}, P_{K_\Gamma}Q_{XX'K|K'=k'}, P_{K_\Gamma K'_\Gamma}Q_{XX'KK'}\}$.

Theorem 5. *If* $Q_{X|Y=y,K=k} = 1 \wedge Q_{K|Y=y,X=x} = q_x 2^{-\kappa}$ *then:*

$$Q_{<XX'KK'_\Gamma>} = \begin{cases} \frac{1}{2} & |\Gamma| > 1 \\ 0 & else \end{cases} \tag{7}$$

$$Q_{<XX'_\Lambda KK'>} \le \max_{x\in\Lambda}\{\sqrt{(q_x-1)2^{-\kappa}}, (\sum_{x\in\Lambda} q_x - 1)2^{-\kappa}, \sqrt{(\sum_{x\in\Lambda} q_x - 1)2^{-\kappa|\Lambda|}}\}. \tag{8}$$

Proof. The Eq.(7) can be get directly from Lemma5.

- Since $Q_{K|Y=y,X=x} = q_x 2^{-\kappa}$, for $x \in \Lambda$, the fastest way to get collision is for random select a $k_1, \ldots, k_t$ getting $y = F_{k_i}(x)$, checking $F_{k_i}(x) = F_{k_j}(x)$ equals or not, similar as proof of Lemma3, $Q_{KK'|X=x,x\in\Lambda} = \sqrt{(q_x-1)2^{-\kappa}}$.
- $Q_{X|Y=y,K=k} = 1 \Rightarrow Q_{XK'|K=k} = \frac{1}{2} \Rightarrow Q_{X=x,K'|K=k} = P_{Y|X=x}(F_k(x)) - 1$, we have $P_{X_\Lambda}Q_{XK'|K} = (\sum\limits_{x\in\Lambda} q_x - 1)2^{-\kappa}$.
- If $|\Lambda| = 1$, $Q_{KX=x,X'=x'} = 0$, or else for selected x, x' searching k satisfying $F_k(x) = F_k(x')$, the success firstly needs exist of k satisfying $F_k(x) = F_k(x')$ for selected x, x', since $Q_{K|Y=y,X=x} \le q_x 2^{-\kappa}$ then $Q_{KX=xX'=x'} \le \#\{F_k(x) = F_k(x') | x, x' \in \Lambda, k \in \{0,1\}^\kappa\}2^{-\kappa}$.

– if $|\Lambda| = 1$, $Q_{K,K'|X=x,X'=x';x,x'\in\Lambda} = 0$, or else for given x, x' the fastest way to find k, k' is random select $k_1, k_2, ..$, compute $y_i = F_{k_i}(x)$,$y'_i = F_{k_i}(x')$ then check y_i equals y'_j or not, since $Q_{K|Y=y,X=x} \leq q_x 2^{-\kappa}$, from Lemma3 we get $P_{X_\Lambda X'_\Lambda} Q_{XX'KK'} = \sqrt{(\sum_{x\in\Lambda} q_x - 1)2^{-\kappa|\Lambda|}}$. □

Theorem 6. *If* $Q_{K|Y=y,X=x} = 1 \wedge Q_{X|Y=y,K=k} = q_k 2^{-\iota}$ *then:*

$$Q_{<XX'KK'_\Gamma>} \leq \max_{k\in\Gamma}\{\sqrt{(q_k-1)2^{-\iota}}, (\sum_{k\in\Gamma} q_k - 1)2^{-\iota}, \sqrt{(\sum_{k\in\Gamma} q_k - 1)2^{-\iota|\Gamma|}}\}. \quad (9)$$

$$Q_{<XX'_\Lambda KK'>} = \frac{1}{2}. \quad (10)$$

4 Hash Properties of Iterated Structure

In this section, we give the proves of that if the compression function is free start preimage resistant and collision resistant, then the hash function is free start preimage resistant and but not free start collision resistant, if the compression function is fix start collision resistant and preimage resistant then the hash function is fix start collision resistant and preimage resistant, and also the upper bounds of collision resistance and preimage resistance are given based on the condition probabilities $P_{Y|X=x}(y)$ and $P_{Y|K=k}(y)$ of compression function F. And also if the compression function is not immune to free start preimage resistant, then the compression function should be designed with minimum values of $P_{Y|X=x}(y)$ and $P_{Y|K=k}(y)$, which imply the best design require the Y is uniformly distributed in $\{0,1\}^n$ for each k and each x, if $n = \kappa = m$ then the best design of compression function is permutation for each k and each x.

Lemma 10. *Let* $F : \{0,1\}^n \times \{0,1\}^\kappa \to \{0,1\}^n$, $y = F_k(x)$, $y \in \{0,1\}^n$, $x \in \{0,1\}^n$, $H_X : \{0,1\}^n \times \{0,1\}^{\kappa\cdot t} \to \{0,1\}^n$, $z = F_{m_t}(\ldots F_{m_1}(x)\ldots)$, $z \in \{0,1\}^n$, $m = m_t\|\ldots\|m_1 \in \{0,1\}^{\kappa\cdot t}$ *and* $m_1, \ldots, m_t$ *are independent from each other then:*

– $P_{Z|M=m}(z) \leq {q_k}^t 2^{-n}$
– $P_{Z|X=x}(z) \leq q_x 2^{-\kappa}$.

Proof. It is clear $t = 1$ the inequality is correct, when $t = 2$:

$$\begin{aligned} & P_{Z|M=m}(z) = P_{Z|M=m_2\|m_1}(z) \\ &= \sum_x P_X(x) P_{Z|M=m_2\|m_1,X=x}(z = F_{m_2}(F_{m_1}(x))) \\ &= \sum_x \sum_u P_X(x) P_{Z|M=m_2\|m_1,X=x}(z = F_{m_2}(u), u = F_{m_1}(x)) \end{aligned}$$

$$= \sum_u P_{Z|M_2=m_2,U=u}(z = F_{m_2}(u)) \sum_x P_X(x) P_{U|M_1=m_1,X=x}(u = F_{m_1}(x))$$
$$= \sum_u P_{Z|M_2=m_2,U=u}(z = F_{m_2}(u)) P_{U|M_1=m_1}(u)$$
$$\leq q_k \sum_u \frac{1}{2^n} P_{Z|M_2=m_2,U=u}(z = F_{m_2}(u)) \leq q_k P_{Z|M_2=m_2}(z)$$

$$P_{Z|X=x}(z) = \sum_{m_1,m_2} P_M(m_1) P_M(m_2) P_{Z|M=m_2\|m_1,X=x}(z = F_{m_2}(F_{m_1}(x)))$$
$$= \sum_{m_1,m_2} \sum_u P_M(m_1) P_M(m_2) P_{Z|M=m_2\|m_1,X=x}(z = F_{m_2}(u), u = F_{m_1}(x))$$
$$= \sum_{m_2} \sum_u P_M(m_2) P_{Z|M_2,U=u}(z = F_{m_2}(u)) \sum_{m_1} P_M(m_1) P_{U|M_1,X}(u = F_{m_1}(x))$$
$$= \sum_{m_2} \sum_u P_M(m_2) P_{Z|M_2,U=u}(z = F_{m_2}(u)) P_{U|X=x}(u)$$
$$= \sum_u P_{Z|U=u}(z) P_{U|X=x}(u) \leq q_x 2^{-\kappa} \sum_u P_{U|X=x}(u) = q_x/2^{\kappa}.$$

Let assume when $t \leq l-1$ the inequality is true, when $t = l$

$$P_{Z|M=m}(z) = \sum_x P_X(x) P_{Z|M'=m'\|m_1,X=x}(z = H_X(m', F_{m_1}(x)))$$
$$= \sum_u P_{Z|M'=m',U=u}(z = H_X(m', u)) P_{U|M_1=m_1}(u)$$
$$\leq q_k \sum_u \frac{1}{2^n} P_{Z|M'=m',U=u}(z = H_X(m', u)) \leq {q_k}^l 2^{-n}$$

$$P_{Z|X=x}(z) = \sum_{m',m_1} P_M(m') P_M(m_1) P_{Z|M=m'\|m_1,X=x}(z = H_X(m', (F_{m_1}(x)))$$
$$= \sum_{m',m_1,u} P_{M'}(m') P_M(m_1) P_{Z|M=m'\|m_1,X=x,U=u}(z = H_X(m', u), u = F_{m_1}(x))$$
$$= \sum_{m'} \sum_u P_{M'}(m') P_{Z|M'=m',U=u}(z = H(m', u)) P_{U|X=x}(u)$$
$$= \sum_u P_{Z|U=u}(z) P_{U|X=x}(u) \leq q_x 2^{-\kappa} \sum_u P_{U|X=x}(u) = q_x 2^{-\kappa}.$$

From induction principle we get the conclusions. □

Lemma 11. *Let* $F : \{0,1\}^{\iota} \times \{0,1\}^n \to \{0,1\}^n$, $H_K : \{0,1\}^{\iota \cdot t} \times \{0,1\}^n \to \{0,1\}^n, k \in \{0,1\}^n, m \in \{0,1\}^{\iota \cdot t}, y \in \{0,1\}^n$, $z \in \{0,1\}^n$, $y = F_k(x)$, $z = F_{\ldots(F_{F_k(m_1)}(m_2))\ldots}(m_t)$ *and* $m_1, \ldots, m_t$ *are independent from each other then:*

- $P_{Z|M=m}(z) \leq \max_x {q_x}^t 2^{-n}$

– $P_{Z|K=k}(z) \leq \max_k q_k 2^{-\iota}$.

Proof. It is clear $t = 1$ the inequations are correct. Let assume when $t \leq l-1$ the inequations are correct.

$$P_{Z|M=m}(z) = \sum_k P_K(k) P_{Z|M'=m'\|m_1,K=k}(z = H_K(m', F_k(m_1)))$$
$$= \sum_u P_{Z|M'=m',U=u}(z = H_K(m', u)) P_{U|M_1=m_1}(u)$$
$$\leq q_x \sum_u \frac{1}{2^n} P_{Z|M'=m',U=u}(z = H_X(m', u)) \leq {q_x}^l 2^{-n}$$

$$P_{Z|K=k}(z) = \sum_{m',m_1} P_M(m') P_M(m_1) P_{Z|M=m'\|m_1,K=k}(z = H_K(m', (F_k(m_1)))$$
$$= \sum_{m',m_1,u} P_{M'}(m') P_M(m_1) P_{Z|M=m'\|m_1,K=k,U=u}(z = H_K(m', u), u = F_k(m_1))$$
$$= \sum_{m'} \sum_u P_{M'}(m') P_{Z|M'=m',U=u}(z = H_K(m', u)) P_{U|K=k}(u)$$
$$= \sum_u P_{Z|U=u}(z) P_{U|K=k}(u) \leq q_k 2^{-\kappa} \sum_u P_{U|K=k}(u) = q_k 2^{-\iota}.$$

From induction principle we get the conclusions. □

Theorem 7. *If $F : \{0,1\}^n \times \{0,1\}^\kappa \to \{0,1\}^n$ is preimage resistant and collision resistant, $H_X : \{0,1\}^n \times \{0,1\}^{\kappa \cdot t} \to \{0,1\}^n, x \in \{0,1\}^n, m \in \{0,1\}^{\kappa \cdot t}, y \in \{0,1\}^n$, $z \in \{0,1\}^n$, $y = F_k(x)$ and $z = F_{m_t}(\ldots F_{m_1}(x)\ldots)$ then:*

$$Q_{[MX|Z]} \leq 2 \max_{x,k}\{q_x 2^{-\kappa}, q_k 2^{-n}\} \tag{11}$$

$$Q_{<MM'XX'>} = 1/2 \tag{12}$$

Proof.

– Let assume for given y we find m, x satisfying $H_X(m_t\| \ldots \|m_1, x) = y$ then we find $H_X(m_{t-1}\| \ldots \|m_1, x), m_t$ satisfying $F_{m_t}(H_X(m_{t-1}\| \ldots \|m_1, x)) = y$, which implies $Q_{[XK|Y]} \geq \frac{1}{\frac{1}{Q_{[MX|Z]}}+t} \geq \frac{Q_{[MX|Z]}}{2}$, from Theorem1 we get the conclusion.
– We get $H_X(m_2\|m_1, x) = H_X(m_2, H_X(m_1, x))$, then we find collision. □

Theorem 8. *If $F : \{0,1\}^\iota \times \{0,1\}^n \to \{0,1\}^n$, is preimage resistant and collision resistant, $H_K : \{0,1\}^{\iota \cdot t} \times \{0,1\}^n \to \{0,1\}^n, k \in \{0,1\}^n, m \in \{0,1\}^{\iota \cdot t}, y \in \{0,1\}^n$, $z \in \{0,1\}^n$, $y = F_k(x)$ and $z = F_{\ldots(F_{F_k(m_1)}(m_2))\ldots}(m_t)$ then:*

$$Q_{[KM|Z]} \leq 2 \max_{x,k}\{q_x 2^{-\kappa}, q_k 2^{-n}\} \tag{13}$$

$$Q_{<KK'MM'>} = 1/2 \tag{14}$$

Theorem 9. *If* $F : \{0,1\}^n \times \{0,1\}^\kappa \to \{0,1\}^n$, $H_X : \{0,1\}^n \times \{0,1\}^{\kappa\cdot *} \to \{0,1\}^n, x \in \{0,1\}^n, m \in \{0,1\}^{\kappa\cdot *}, y \in \{0,1\}^n$, $z \in \{0,1\}^n$, $y = F_k(x)$, $z = F_{m_*}(\ldots F_{m_1}(x)\ldots)$, $m_1, \ldots, m_*$ *are independent from each other,*

- *if* $Q_{K|Y=y,X=x} = 1 \wedge Q_{X|Y=y,K=k} = q_k 2^{-\iota}$ *then*

$$Q_{[X_\Lambda M|Z]} = \frac{\kappa}{|M|}, \qquad Q_{<XX'_\Lambda MM'>} = \frac{\kappa}{|M|+|M'|}$$

- *if* $Q_{X|Y=y,K=k} = 1 \wedge Q_{K|Y=y,X=x} = q_x 2^{-\kappa}$, $\Lambda' \stackrel{\triangle}{=} \{H_X(m',x), x \in \Lambda\} \cup \Lambda$ *then:*

$$Q_{[X_\Lambda M|Z]} = \max\{\sum_{x\in\Lambda} q_x 2^{-\kappa}, q_k^{\frac{|M|}{\kappa}} 2^{-n}\} \tag{15}$$

$$Q_{<XX'_\Lambda MM'>} \le \max_{x\in\Lambda,k}\{\frac{q_k^{\frac{|M|}{\kappa}}}{2^n}, \sqrt{\frac{(q_x-1)}{2^\kappa}}, \frac{\sum_{x\in\Lambda} q_x - 1}{2^\kappa}, \sqrt{\frac{\sum_{x\in\Lambda'} q_x - 1}{2^{\kappa|\Lambda'|}}}\} \tag{16}$$

Proof.

- The conclusions $Q_{[X_\Lambda M|Y]} = \frac{\kappa}{|M|}$, $Q_{<XX'_\Lambda MM'>} = \frac{\kappa}{|M|+|M'|}$ can be get by the direct computation, since $Q_{K|X_\Lambda Y} = 1$.
- there are two ways to find the preimage:
  - Case 1 : Finding the preimage of z, directly searching $m \in \{0,1\}^{\kappa\cdot *}$ satisfying $z = H_X(m,x)$ where $x \in \Lambda$. $Q_{K|Y=y,X=x} = q_x 2^{-\kappa}$ implies for given z, x the only way of finding m satisfying $z = H_X(m,x)$ is exhaustive search, more precisely, $Q_{M|Z=z,X=x} = q_x 2^{-\kappa}\frac{\kappa}{|M|}$. From Lemma10 and Theorem3 we get $Q_{[M|X_\Lambda Z]} = \sum_{x\in\Lambda} q_x \kappa 2^{-\kappa}$.
  - Case 2 : for given z, search $m' \in \{0,1\}^{\kappa\cdot t'}$, $m'' \in \{0,1\}^{\kappa\cdot t''}$, satisfying $z = H_X(m'',u)$ and $u = H_X(m',x)$ where $x \in \Lambda$:
    - Select m' randomly, searching m'', let $\Lambda' \stackrel{\triangle}{=} \{H_X(m',x), x \in \Lambda\}$, the problem become case 1;
    - Select m'' randomly, get u from $z = H_X(m'',u)$, then searching m' satisfying $u = H_X(m',x)$, equals finding the preimage of u;
    - Guessing m' and m'', compute u and u' from $u = H_X(m',x)$ and $z = H_X(m'',u')$, let $t = |m''|$, the probability of $u = u'$ smaller than $\max\{q_k^t 2^{-n}, q_x 2^{-\kappa}\}$, more precisely, if the compression function is designed with property of that, $\exists \dot{z} \in \{0,1\}^n$, $\dot{m} \in \{0,1\}^{\kappa t}$ satisfy $P_{Z|M=\dot{m}}(\dot{z}) = q_k^t$ and $q_k > 1$, then the complexity of finding preimage of $\dot{z}$ is $\frac{q_k^t}{2^n}$, where we search m satisfy $\dot{z} = H_X(\dot{m}\|m,x)$.

  From Case 1 and Case 2, we get $Q_{[X_\Lambda M|Z]} = \max\{\sum_{x\in\Lambda} q_x 2^{-\kappa}, q_k^{\frac{|M|}{\kappa}} 2^{-\iota}\}$.
- there are three ways to find the collision,let :

- Case 1: Finding collision of H_X means searching $m' \in \{0,1\}^{\kappa \cdot t'}$, $m'' \in \{0,1\}^{\kappa \cdot t''}$ satisfying $H_X(m',x) = H_X(m'',x)$ with $x \in \Lambda$. $Q_{K|Y=y,X=x} = q_x 2^{-\kappa}$ implies for given z, x the only way of finding m satisfying $z = H_X(m,x)$ is exhaustive search, more precisely, $Q_{M|Z=z,X=x} = q_x 2^{-\kappa} \frac{\kappa}{|M|}$. From Lemma10 and Theorem5 we get the conclusion: $Q_{<XX'_\Lambda MM'>} \leq \max\limits_{x\in\Lambda}\{\sqrt{(q_x-1)2^{-\kappa}}, (\sum\limits_{x\in\Lambda} q_x - 1)2^{-\kappa}, \sqrt{(\sum\limits_{x\in\Lambda} q_x - 1)2^{-\kappa|\Lambda|}}\}$.
- Case 2: search $m \in \{0,1\}^{\kappa\cdot t}$, $m' \in \{0,1\}^{\kappa\cdot t'}$, $m'' \in \{0,1\}^{\kappa\cdot t''}$, satisfying $H_X(m,x) = H_X(m'',u)$ and $u = H_X(m',x)$ where $x \in \Lambda$:
  - if we randomly select m searching m', m'', the problem becomes finding a primage of $z = H_X(m,x)$;
  - If we randomly select m' get u from $u = H_X(m',x)$, then search m and m'' satisfying $H(m,x) = H_X(m'',u)$, let $\Lambda' \stackrel{\triangle}{=} \{H_X(m',x), x \in \Lambda\} \cup \Lambda$, the problem become case 1 where $x \in \Lambda'$;
  - If we randomly select m'' search m, m' checking $H_X(m'', H_X(m',x)) = H_X(m,x)$ being satisfied or not which needs more computation than for given m'' finding z and m' satisfying $z = H_X(m'', H_X(m',x))$.
- Case 3: search $m \in \{0,1\}^{\kappa\cdot t}, m' \in \{0,1\}^{\kappa\cdot t'}, \bar{m} \in \{0,1\}^{\kappa\cdot\bar{t}}, \bar{m}' \in \{0,1\}^{\kappa\cdot\bar{t}'}$ satisfying $H_X(m', H_X(m,x)) = H_X(\bar{m}', H_X(\bar{m},x)$ where $x \in \Lambda$, similar as case 2, case 3 needs more computation than case 2.

From Case 1, Case 2 and Case 3, we get the conclusion: $Q_{<XX'_\Lambda MM'>} \leq \max\limits_{x\in\Lambda,k}\{q_k^{\frac{|M|}{\kappa}} 2^{-n}, \sqrt{(q_x-1)2^{-\kappa}}, (\sum\limits_{x\in\Lambda} q_x - 1)2^{-\kappa}, \sqrt{(\sum\limits_{x\in\Lambda} q_x - 1)2^{-\kappa|\Lambda|}}\}$. □

Theorem 10. *Let* $F : \{0,1\}^n \times \{0,1\}^\kappa \to \{0,1\}^n$, $H_K : \{0,1\}^{\iota\cdot t} \times \{0,1\}^n \to \{0,1\}^n, k \in \{0,1\}^n, m \in \{0,1\}^{\iota\cdot t}, y \in \{0,1\}^n$, $z \in \{0,1\}^n$, $y = F_k(x)$, $z = F_{\ldots(F_{F_k(m_1)}(m_2))\ldots}(m_t)$ *and* $m_1, \ldots, m_t$ *are independent from each other then:*

- *if* $Q_{X|Y=y,K=k} = 1 \wedge Q_{K|Y=y,X=x} \leq q_x 2^{-n}$ *then*

$$Q_{[K_\Gamma M|Z]} = \frac{\iota}{|M|}, \qquad Q_{<KK'_\Gamma MM'>} = \frac{\iota}{|M| + |M'|}$$

- *if* $Q_{K|Y=y,X=x} = 1 \wedge Q_{K|Y=y,K=k} \leq q_x 2^{-\iota}$, $\Gamma' \stackrel{\triangle}{=} \{H_X(m',x), x \in \Gamma\} \cup \Gamma$ *then:*

$$Q_{[K_\Gamma M|Z]} = \max\{\sum_{k\in\Gamma} q_k 2^{-\kappa}, q_x^{\frac{|M|}{\kappa}} 2^{-n}\} \tag{17}$$

$$Q_{<KK'_\Gamma MM'>} \leq \max_{k\in\Gamma,x}\{\frac{q_x^{\frac{|M|}{\kappa}}}{2^n}, \sqrt{\frac{(q_k-1)}{2^\kappa}}, \frac{\sum\limits_{k\in\Gamma} q_k - 1}{2^\kappa}, \sqrt{\frac{\sum\limits_{k\in\Gamma'} q_k - 1}{2^{\kappa|\Gamma'|}}}\} \tag{18}$$

Theorem 11. *Let* $F : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^n$, $H_K : \{0,1\}^{n\cdot t} \times \{0,1\}^n \to \{0,1\}^n$, $H_X : \{0,1\}^{n\cdot t} \times \{0,1\}^n \to \{0,1\}^n$, $m = m_t \| \ldots \| m_t$, $m_1, \ldots, m_t$ *are*

independent from each other. If $Q_{K|Y=y,K=k} \leq q_k 2^{-n} \wedge Q_{K|Y=y,X=x} \leq q_x 2^{-n}$, $\Lambda' \stackrel{\triangle}{=} \{H_X(m',x), x \in \Lambda\} \cup \Lambda$,$\Gamma' \stackrel{\triangle}{=} \{H_X(m',x), x \in \Gamma\} \cup \Gamma$ *then:*

$$Q_{[X_\Lambda M|Z]} = \max_{x\in\Lambda}\{q_x 2^{-n}, q_k^{\frac{|M|}{n}} 2^{-n}\} \tag{19}$$

$$Q_{<XX'_\Lambda MM'>} \leq \max_{x\in\Lambda,k}\{\frac{q_k^{\frac{|M|}{n}}}{2^n}, \sqrt{\frac{(q_x-1)}{2^n}}, \sqrt{\frac{\sum\limits_{x\in\Lambda'} q_x - 1}{2^{n|\Lambda'|}}}\} \tag{20}$$

$$Q_{[K_\Gamma M|Z]} = \max_{k\in\Gamma}\{q_k 2^{-n}, q_x^{\frac{|M|}{n}} 2^{-n}\} \tag{21}$$

$$Q_{<KK'_\Gamma MM'>} \leq \max_{k\in\Gamma,x}\{\frac{q_x^{\frac{|M|}{n}}}{2^n}, \sqrt{\frac{(q_k-1)}{2^n}}, \sqrt{\frac{\sum\limits_{k\in\Gamma'} q_k - 1}{2^{n|\Gamma'|}}}\} \tag{22}$$

Theorem11 tell us on condition of the compression function F is free start preimage resistant and free start collision resistant, the best design of H_X and H_K have properties of $q_k = 1$ and $q_x = 1$.

5 Collision Resistance of PGV Schemes

We assume block cipher $E : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^n$ has no weakness, that means for given y, x no ways to finding k except exhaustive search, we also assume $P_{Y|X=x}(y) = P_{Y|K=k}(y) = \frac{1}{2^n}$, the block cipher E with perfect security and perfect key distribution. The security of 24 PGV schemes is summarized in tables 1 where we give up to consider the constant value v, in black box model the value v does not influence the security of compression function and hash function. The functions are numbered in BRS[12], where the $F_1 \sim F_{12}$ are the group one schemes which is immune to free start collision resistance and $F_{13} \sim F_{20}$ are the group two schemes which are not immune to free start collision resistance and immune to fix start collision resistance. In fact those 24 schemes are derive from 12 compression function with different fix start and four of which are not immune to fix start collision resistance.

Theorem 12. *If block cipher E is a random oracle model with perfect security and perfect key distribution, $P_{Y|K=k}(y), P_{Y|X=x}(y)$ of F_i $1 \leq i \leq 24$ are equal to that of E.*

Proof. Since $y = E_k(x)$ is random orale, x,k and $E_k(x)$ are independent from each other. We give the prove of the most famous mode $y' = E_k(x) \oplus x$.

$$P_{Y|K=k}(y) = \sum_x P_X(x) P_{Y|X=x,K=k}(y = E_k(x) \oplus x)$$

$$= \sum_{x,x'} P_X(x) P_{X'}(x') P_{Y|X=x,K=k,X'=x'}(y = E_k(x) \oplus x')$$

$$= \sum_{x,x'} \sum_{t} P_X(x) P_{Y|X=x,T=t}(y = x \oplus t) P_{X'}(x') P_{T|X'=x',K=k}(t)$$
$$= \sum_{x} \sum_{t} P_X(x) P_{Y|X=x,T=t}(y = x \oplus t) P_{T|K=k}(t)$$
$$= p_k \sum_{x} \sum_{t} P_X(x) P_{Y|X=x,T=t}(y = x \oplus t) = p_k$$

$$P_{Y|X=x}(y) = \sum_{k} P_K(k) P_{Y|X=x,K=k}(y = E_k(x) \oplus x)$$
$$= \sum_{k} \sum_{t} P_K(k) P_{Y|X=x,K=k}(y = x' \oplus t; t = E_k(x))$$
$$= \sum_{k} \sum_{t} P_K(k) P_{Y|X=x,T=t}(y = x \oplus t) P_{T|X=x,K=k}(t)$$
$$= \sum_{t} P_{Y|X=x,T=t}(y = x \oplus t) p_x = p_x$$

The prove of mode 13 is:

$$P_{Y|K=k}(y) = \sum_{x} P_X(x) P_{Y|X=x,K=k}(y = E_{k \oplus x}(x'))$$
$$= \sum_{x,t} P_X(x) P_{Y|X=x,K=k}(t = k \oplus x; y = E_t(x'))$$
$$= p_k \sum t P_{Y|X=x,K=k}(t = k \oplus x) = p_k$$

$$P_{Y|X=x}(y) = \sum_{k} P_K(k) P_{Y|X=x,K=k}(y = E_{k \oplus x}(x))$$
$$= \sum_{k,t} P_K(k) P_{Y|X=x,K=k}(t = k \oplus x; y = E_t(x')) = p_x$$

other modes can be proved in similar way. □

Theorem 13. *$F_i, 1 \leq i \leq 12$ are free start preimage resistant and free start collision resistant.*

Theorem 14. *$F_i, 1 \leq 13 \leq 24$ are not free start preimage resistant and not free start collision resistant.*

Proof.

- F_{13}:$\forall y, k$ compute $x = E_k^{-1}(y)$, let $k' = x \oplus k$ then $E_{k' \oplus x}(x) = y$;
- F_{14}:$\forall y, k$ compute $x = E_k^{-1}(y \oplus k)$, let $k' = x \oplus k$ then $E_{k' \oplus x}(x) \oplus k = y$;
- F_{15}:$\forall y, k$ compute $x = E_k^{-1}(y \oplus k)$, let $k' = x, x' = k$ then $E_{k'}(x') = y$;
- F_{16}:$\forall y, k$ compute $x = E_k^{-1}(y)$, let $k' = x, x' = k \oplus x$ then $E_{k' \oplus x'}(k') = y$;
- F_{17}:$\forall y, k$ compute $x = E_k^{-1}(y \oplus k)$, let $k' = x, x' = k$ then $E_{x'}(k') \oplus x' = y$;

- F_{18}:$\forall y, k$ compute $x = E_k^{-1}(y \oplus k)$, let $k' = x, x' = k \oplus x$ then $E_{x' \oplus k'}(k') \oplus x' \oplus k' = y$;
- F_{19}:$\forall y, k$ compute $x = E_k^{-1}(y)$, let $k' = x \oplus k, x' = k$ then $E_{x'}(x' \oplus k) = y$;
- F_{20}:$\forall y, k$ compute $x = E_k^{-1}(y \oplus k)$, let $k' = x \oplus k$ then $E_{k' \oplus x}(x) \oplus k = y$;
- F_{21}:$\forall y, k$ compute $x = E_k^{-1}(y)$ then $E_k(x) = y$;
- F_{22}:$\forall y, k$ compute $x = E_k^{-1}(y \oplus k)$ then $E_k(x) \oplus k = y$;
- F_{23}:$\forall y, k$ compute $x = E_k^{-1}(y)$, let $x' = x \oplus k$ then $E_k(x' \oplus k) = y$;
- F_{24}:$\forall y, k$ compute $x = E_k^{-1}(y \oplus k)$, let $x' = k \oplus k$ then $E_k(x' \oplus k) \oplus k = y$. □

Theorem 15. *If block cipher E is a random oracle model with perfect security and perfect key distribution the hash functions H_X are fix start collision resistant and fix start preimage resistant where compression functions are $F_i, i \in \{5 \sim 8, 10, 12, 15 \sim 20\}$ and H_K are fix start preimage resistant and fix start collision resistant where compression functions are $F_i, i \in \{1 \sim 4, 9, 11, 13, 14\}$.*

The theorem12 needs block cipher is random oracle, or else (x and k are not independent from $E_k(x)$. In fact, E is not a random oracle and only F_{15}, F_{21} with properties of $P_{Y|X=x} = P_{Y|K=k} = \frac{1}{2^n}$, which implies the best compression functions is F_{15}. So we have conclusion that if E is designed to be $P_{Y|X=x} = P_{Y|K=k} = \frac{1}{2^n}$, then the block cipher can be used to construct a secure hash function.

6 Conclusion

The main conclusion of this paper is that if no way to design the compression $F(k, x)$ immune to free start preimage resistant, then the best design of compression function is a block cipher with perfect key distribution and perfect security where the hash function has M-D structure. So the design of block cipher and hash function can be one problem and the design of key schedule algorithm of block cipher become important than before.

References

1. B.Preneel: The State of Cryptographic Hash Functions. In Lectures on Data Security, Lecture Notes in Computer Science, Vol. 1561. Springer-Verlag, Berlin Heidelberg New York (1999) 158-182.
2. B. Preneel, R. Govaerts, and J. Vandewalle, " Hash functions based on block ciphers,", In Advances in Cryptology -CRYPTO'93, Lecture Notes in Computer Science,pages 368-378. Springer-Verlag, 1994.
3. B.Preneel, V. Rijmen, A.Bosselaers: Recent Developments in the Design of Conventional Cryptographic Algorithms. In State of the Art and Evolution of Computer Security and Industrial Cryptography. Lecture Notes in Computer Science, Vol 1528. Springer-Verlag, Berlin Heidelberg New York(1998) 106-131.
4. B. Van Rompay,Analysis and design of cryptographic hash functions, MAC algorithms and block cipher, K. U. Leuven, Juni 2004
5. C.Chchin. Entropy Measures and Uncoditional Security in Cryptography, PHD thesis.

Table 1. Summary of results. Columen1 is the number of hash function which is given by BRS[12]. Column2 are the compression functions to build hash function. Column3,4,5 are the probabilities of finding the preimages in one time computation of F. Column 6 are the hash function and column 7 are the compression function F_i. We write $w_i = m_i \oplus h_{i-1}$.

i	$y =$	$Q_{K\|XY}$	$Q_{X\|KY}$	$Q_{XK\|Y}$	$z =$	$h_i =$
15	$E_k(x)$	$p_k{}^{\frac{1}{2}}$	1	1	H_X	$E_{m_i}(h_{i-1}) \oplus v$
21	$E_k(x)$				H_K	$E_{h_{i-1}}(m_i) \oplus v$
19	$E_k(x \oplus k)$	$p_k{}^{\frac{1}{2}}$	1	1	H_X	$E_{m_i}(w_i) \oplus v$
23	$E_k(x \oplus k)$				H_K	$E_{h_{i-1}}(w_i) \oplus v$
5	$E_k(x) \oplus x$	$p_k{}^{\frac{1}{2}}$	$p_x{}^{\frac{1}{2}}$	$p^{\frac{1}{2}}$	H_X	$E_{m_i}(h_{i-1}) \oplus h_{i-1}$
1	$E_k(x) \oplus x$				H_K	$E_{h_{i-1}}(m_i) \oplus m_i$
17	$E_k(x) \oplus k$	$p_k{}^{\frac{1}{2}}$	1	1	H_X	$E_{m_i}(h_{i-1}) \oplus m_i$
22	$E_k(x) \oplus k$				H_K	$E_{h_{i-1}}(m_i) \oplus h_{i-1}$
7	$E_k(x) \oplus x \oplus k$	$p_k{}^{\frac{1}{2}}$	$p_x{}^{\frac{1}{2}}$	$p^{\frac{1}{2}}$	H_X	$E_{m_i}(h_{i-1}) \oplus w_i$
3	$E_k(x) \oplus x \oplus k$				H_K	$E_{h_{i-1}}(m_i) \oplus w_i$
8	$E_k(x \oplus k) \oplus x$	$p_k{}^{\frac{1}{2}}$	$p_x{}^{\frac{1}{2}}$	$p^{\frac{1}{2}}$	H_X	$E_{m_i}(w_i) \oplus h_{i-1}$
4	$E_k(x \oplus k) \oplus x$				H_K	$E_{h_{i-1}}(w_i) \oplus m_i$
20	$E_k(x \oplus k) \oplus k$	$p_k{}^{\frac{1}{2}}$	1	1	H_X	$E_{m_i}(w_i) \oplus m_i$
24	$E_k(x \oplus k) \oplus k$				H_K	$E_{h_{i-1}}(w_i) \oplus h_{i-1}$
6	$E_k(k \oplus x) \oplus x \oplus k$	$p_k{}^{\frac{1}{2}}$	$p_x{}^{\frac{1}{2}}$	$p^{\frac{1}{2}}$	H_X	$E_{m_i}(w_i) \oplus w_i$
2	$E_k(k \oplus x) \oplus x \oplus k$				H_K	$E_{h_{i-1}}(w_i) \oplus w_i$
16	$E_{k\oplus x}(x)$	$p_k{}^{\frac{1}{2}}$	$p_x{}^{\frac{1}{2}}$	1	H_X	$E_{w_i}(h_{i-1}) \oplus v$
13	$E_{k\oplus x}(x)$				H_K	$E_{w_i}(m_i) \oplus v$
10	$E_{k\oplus x}(x) \oplus x$	$p_k{}^{\frac{1}{2}}$	$p_x{}^{\frac{1}{2}}$	$p^{\frac{1}{2}}$	H_X	$E_{w_i}(h_{i-1}) \oplus h_{i-1}$
9	$E_{k\oplus x}(x) \oplus x$				H_K	$E_{w_i}(m_i) \oplus m_i$
12	$E_{k\oplus x}(x) \oplus k$	$p_k{}^{\frac{1}{2}}$	$p_x{}^{\frac{1}{2}}$	$p^{\frac{1}{2}}$	H_X	$E_{w_i}(h_{i-1}) \oplus m_i$
11	$E_{k\oplus x}(x) \oplus k$				H_K	$E_{w_i}(m_i) \oplus h_{i-1}$
18	$E_{k\oplus x}(x) \oplus x \oplus k$	$p_k{}^{\frac{1}{2}}$	$p_x{}^{\frac{1}{2}}$	$p^{\frac{1}{2}}$	H_X	$E_{w_i}(h_{i-1}) \oplus w_i$
14	$E_{k\oplus x}(x) \oplus x \oplus k$				H_K	$E_{w_i}(m_i) \oplus w_i$

6. C.E. Shannon. "Communication theory of secrecy systems,", Bell System Technical Journal, 28:656 C 715, 1949.
7. C. H. Meyer and S. M. Matyas. Cryptography: a New Dimension in Data Security. Wiley & Sons, 1982.
8. E.Biham and R.Chen. Near-Collisions of SHA-0,In Advances in Cryptology CRYPTO'2004, LNCS 3152,pp290-305,2004.
9. E.Biham and R.Chen. Near-Collisions of SHA-0 and SHA-1. In Selected Areas in Cryptography-SAC 2004.
10. M. O. Rabin. Digitalized Signatures. In R. A. Demillo, D. P. Dopkin, A. K. Jones, and R. J. Lipton, editors, Foundations of Secure Computation, pages 155-166, New York, 1978. Academic Press.
11. I.Damgård. A design principle for hash functions. In G. Brassard, editor, Advances in Cryptology-CRYPTO' 89, volume 435 of Lecture Notes in Computer Science. Springer-Verlag, 1990.
12. J.Black, P.Rogaway, and T.Shrimpton, "Black-box analysis of the block-cipher-based hashfunction constructions from PGV". In Advances in Cryptology - CRYPTO'02, volume 2442 of Lecture Notes in Computer Science. Springer-Verlag, 2002.pp.320-335.
13. J. Daemen and V. Rijmen: The Design of Rijndael: AES The Advanced Encryption Standard. Springer, 2002.
14. X. Wang, H. Yu, How to Break MD5 and Other Hash Functions, EUROCRYPT'2005, Springer-Verlag, LNCS 3494, pp19-35, 2005.
15. X. Lai and J. L. Massey: Hash functions based on block ciphers. In Advances in Cryptology Eurocrypt'92, Lecture Notes in Computer Science, Vol. 658. Springer-Verlag, Berlin Hei-delberg New York (1993) 55-70.
16. X. Wang, X. Lai, D.Feng and H.Yu., Cryptanalysis of the Hash Functions MD4 and RIPEMD, EUROCRYPT 2005, Springer-Verlag,LNCS 3494, pp1-18, 2005.